Gradient-Guided Attentional Network for Radio Transient Localization With the Cluster-Feed Telescope

Hailin Cao, Junhui Peng, Jin Fan, Wai Yan Yong, *Graduate Student Member, IEEE*, and Decheng Wu

***Abstract*— Large, single-dish radio telescopes with high sensitivities are ideal for detecting faint radio transients (RTs). However, single-dish radio telescopes possess a limited angular resolution, which limits their accuracy in localizing objects. In this article, we propose to improve the localization accuracy of the RT by exploring the 3-D focal field distributions (3DFFDs) of the dish reflector with a gradient-guided attentional network (GGAN). The LSTM-based attention block of the GGAN achieves the task-oriented adaptive recalibration of 3DFFD features by exploring the significant properties and spatial dependencies of 3DFFD. In addition, a gradient-guided approach is being developed to improve the attention block performance under varying incident angles. The proposed attention mechanism is applied to the convolutional neural network in order to reconstruct 3DFFDs and perceive RT positions based on the reconstructed results. Simulation results indicate that the technique can enable the precise localization of RTs. Moreover, the proposed solution improves the telescope's instantaneous field of view (FOV) compared to a sky survey with the traditional cluster feed telescope.**

***Index Terms*— 3-D focal field distribution (3DFFD), attentional network, localization of radio transients (RTs), single-dish radio telescope, wide field of view (FoV).**

I. Introduction

RADIO transients (RTs) with high localization accuracy can be used as the probes in the exploration of the intergalactic medium and constrain cosmological parameters [1], [2]. Currently, two main types of instruments have been used to detect and locate RTs: interferometers and single-dish radio telescopes. The localization accuracy of the interferometer is directly proportional to the baseline length [3], [4]. As an interferometer used for astronomical observations, it requires a group of reflectors and feeds elements, which generates massive data streams [5]. It poses many challenges in terms of real-time data processing. Alternatively, a single-dish radio telescope has a higher brightness sensitivity than an interferometer with an equal collecting area [6]. The localization precision of the RTs with a single-dish radio telescope increases directly proportionally to the aperture size of the dish [7]. Although increasing the aperture size of the radio telescope dish improves the directivity, resulting in a higher resolution, it reduces the telescope's field of view (FoV). This reveals that the sensing area decreases proportionally as the aperture size increases, resulting in a decrease in the radio telescope's overall survey speed [8]. In addition, expanding the aperture size of the radio telescope is not always possible, especially if the telescope is planned for installation within a restricted space.

A phased array feed (PAF) can be used to improve the FOV and localization accuracy of radio telescopes to solve the aforementioned problems in single-dish radio telescopes. A PAF can generate numerous overlapped beams at the same time utilizing digital beamforming technology. Numerous beams allow for an improvement in the FoV. In addition, the overlapped beams can be used to reduce the uncertainty of RTs' positions [9]. However, PAFs are challenging to design (e.g., impedance matching) and thus costly [10]. Besides that, the cluster feed that employs a cluster of individually feed antennas, such as horn antennas, can be used to generate multiple beams, which also allows for improving the FoV of the radio telescope. The cluster feed systems have been commissioned in some large-aperture single-dish radio telescopes. For instance, the Parkes radio telescope located in Australia and the FAST radio telescope located in China have deployed the cluster feeds to survey the sky [11], [12]. However, these cluster feeds are often electrically large, which results in multibeams in nonoverlapping configurations. In addition, for the cluster feed regimes, the angular resolution of radio telescopes is determined by the beam widths of nonoverlapping beams [7], [8]. Recently, a novel approach for detecting transient radio targets based on the focal-field distribution (FFD) feature matching has been proposed [13]. This article proposed searching for astronomical transients based on variations in the FFDs with different frequencies and incident angles. The main advantage of this technology is

Manuscript received September 28, 2021; revised March 20, 2022; accepted April 11, 2022. Date of publication April 18, 2022; date of current version May 9, 2022. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 51877015, Grant U1831117, Grant U1931129, Grant U20A20157, Grant 11403054, and Grant 11611130023; in part by the Natural Science Foundation of Chongqing under Grant cstc2021jcyj-bsh0198; and in part by the Fundamental Research Funds for the Central Universities under Project 106112017CDJXSYY0002. *(Corresponding author: Jin Fan.)*

Hailin Cao and Junhui Peng are with the Chongqing Key Laboratory of Space Information Network and Intelligent Information Fusion, Chongqing University, Chongqing 400044, China (e-mail: hailincao@cqu.edu.cn; junhuipeng@cqu.edu.cn).

Jin Fan is with the Key Laboratory of Radio Astronomy, National Astronomical Observatories of China, Beijing 100012, China (e-mail: jfan@bao.ac.cn).

Wai Yan Yong is with the Department of Electrical Engineering, University of Twente, 7522 NB Enschede, The Netherlands (e-mail: w.y.yongwaiyan@utwente.nl).

Decheng Wu is with the School of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: wudc@cqupt.edu.cn).

Digital Object Identifier 10.1109/TGRS.2022.3167762

the ability for instantaneous localization based on FFD feature matching without the need for beam scanning. However, the overall performance of this technique is bounded by the number of feed elements. To solve the FFD undersampled problem of the cluster feeds, Wu *et al.* [14] employed a generative adversarial network to reconstruct FFDs at different frequency bands. Then, the deep neural network was utilized to accomplish multifrequency joint estimation of RT positions.

Attention modules in computer vision are employed to enhance the powerful representations of images that are most relevant for a given task. Wang *et al.* [15] used the intermediate features of images to generate the same size attention masks with the residual block as the backbone. The attention masks achieve selective enhancement of the intermediate features in the spatial and channel domains. Based on the intermediate features of images, Woo *et al.* [16] utilized multilayer perceptron (MLP) and convolutional layer to infer the channel and spatial attention masks, respectively, and implemented channel attention and spatial attention sequentially. For the image segmentation task, Fu *et al.* [17] used matrix multiplications to replace the MLP and convolutional layer to achieve attention inference, which helps model the long-range contextual information in segmentation tasks. Focusing on the relationship between channels, Hu *et al.* [18] established the mapping relationships between the channel descriptors of intermediate features and the attention masks to achieve feature recalibration in the channel domain. Recently, it was proposed to use attention blocks in the neural network for multislice 2-D computer tomography (CT) image processing. To differentiate prostate bed from pelvis CT images, Xu *et al.* [19] developed a U-shaped 2-D convolutional network to explore the combination of the center 2-D CT slice and the adjacent slices. Meanwhile, cascaded attention blocks were employed to enhance representative features of the prostate bed. In order to identify the COVID-19 cases from community-acquired pneumonia infections, Ouyang *et al.* [20] applied 3-D convolution operations on CT slice sequence to extract meaningful features. The attention mechanism was adopted in feature mining processing, which compels the model to focus on lesion regions of the lung. A similar concept can be borrowed from the FFD-based RT localization problem. In principle, when RTs come into radio telescopes with various angles, field distributions have off-focus peaks, and thus, the focal plane is not the optimal signal-interception position [21]. Apart from the FFDs in the focal plane, the adjacent slices of FFDs should be taken into consideration.

The analysis of CT slices prompted us to look at both the focal plane and the vicinity plane to get a more comprehensive understanding of RT locations. In addition, a specially designed LSTM-based attention block is utilized to selectively enhance and depress the 3-D focal field distribution (3DFFD) features. Distinct from the previously discussed attention modules, which only consider the intermediate features of the image, the proposed LSTM-based attention block realizes attention not only by utilizing the spatial features but also by considering the spatial dependencies within 3DFFDs. Furthermore, we propose a method of guiding attention block

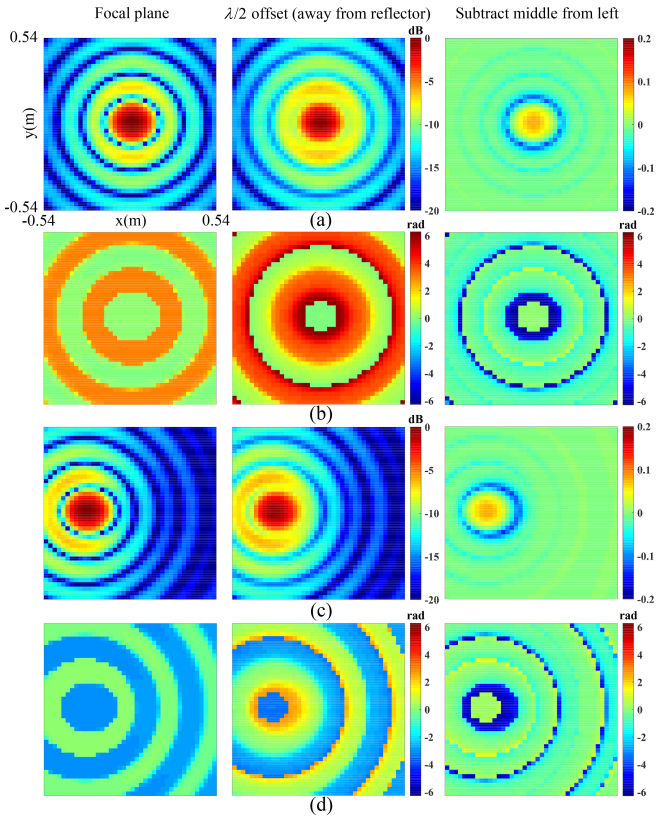

Fig. 1. 3DFFDs on the condition of a 30-m parabolic reflector with a focal length to diameter ratio of 0.467 and linear polarization incident wave with frequency 1.25 GHz. (a) and (c) Amplitude distributions normalized by the intensity peaks. (b) and (d) Phase distributions relative to the phases of the center pixels. In the case of the azimuth angle of 0°, (a)–(d) correspond to zenith angles of 0° and 1°, respectively.

based on the gradient distributions of 3DFFDs. This method provides direct reinforcement of salient features based on prior knowledge without the need for training. The proposed gradient-guided attentional network (GGAN) aims to obtain highly accurate localization of RTs by using two steps. First, the attentional reconstruction network is employed, which permits the reconstruction of 3DFFDs with high resolution from received signals, i.e., undersampled 3DFFDs. Next, the attentional localization network with gradient guidance is utilized to establish mapping relationships between reconstructed 3DFFDs and the RT positions. Our proposed solution allows RT localization to be performed with greater accuracy than the telescope's angular resolution. As well, the proposed solution improves the telescope's instantaneous FOV in comparison with sky survey with cluster feed.

II. 3DFFD Analysis

The xyz coordinate system is established with the focal point as the origin, the focal plane as the xy plane, and the symmetry axis of the reflector as the z-axis. The angles of the incoming wave along the off-z-axis and off-x-axis are defined as zenith and azimuth angles, respectively. As illustrated in Fig. 1, 3DFFDs involve two types of patterns: field amplitude and phase distributions, which were created

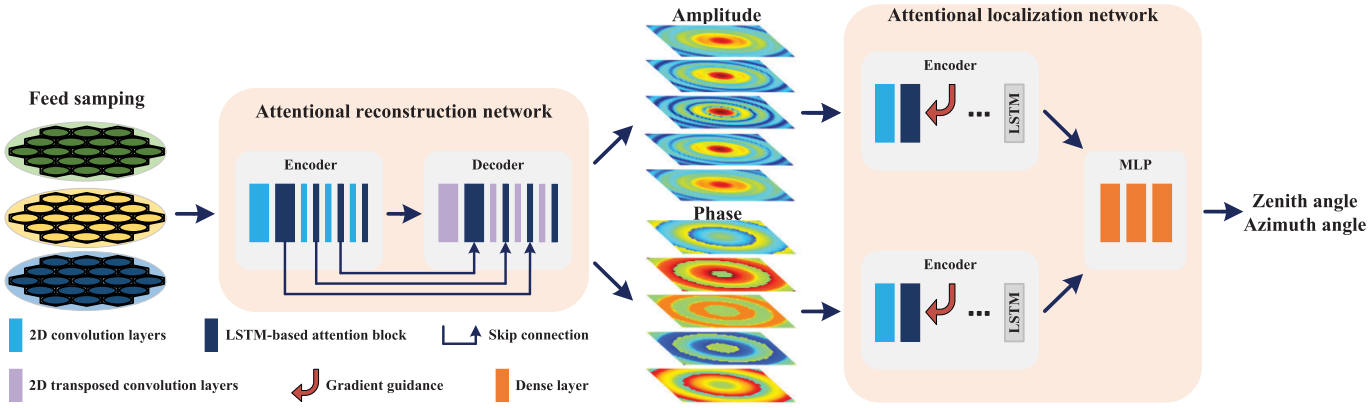

Fig. 2. Illustration of the pipeline of the proposed GGAN. It consists of an attentional reconstruction network and an attentional localization network. The ellipses indicate that previous structures are repeated three times. The inner structure of the LSTM-based attention block and the gradient guidance are illustrated in Figs. 3 and 4, respectively.

using commercial computational electromagnetics software, FEKO. Fig. 1(a) and (c) illustrates the computed amplitude distributions of 3DFFDs. Due to the diffraction effect, the amplitude distribution consists of a bright spot and a series of concentric rings along with the diminishing intensity. Compared with the focal plane, the defocus effect is observed on the amplitude distribution slices on the vicinity planes along the symmetry axis of the reflector, where the intensity peaks diffuse toward distribution gaps. In addition, as can be observed in Fig. 1(b) and (d), the phase distributions are similar to amplitude distributions, and the positions of central regions are consistent for both distributions. However, the phase distributions are relatively flat in circular rings. Besides that, in the $\lambda/2$ offset plane, wave path differences cause variations of phase distribution slices.

By comparing the changes of the 3DFFD caused by the variation of the incident angle in Fig. 1, we observe that axial incident wave produces symmetrical distribution, whereas off-axis incoming wave leads to distortion. Moreover, increasing the zenith angle of RT results in the corresponding movements of central regions along the x-axis. The right column of Fig. 1 shows the differences between distributions on neighborhood slices. As can be seen from the subtraction results, the form and extent of defocusing vary with the incident angle of RT. Likewise, the difference between phase distribution slices varies with the incident angle as well. This suggests that both the spatial features and spatial dependencies of 3DFFDs are linked to RT positions. As a result, when building an attention mask, our attention block takes into account both of the aforementioned parameters, resulting in minimal information loss.

Among these changes of the 3DFFD produced by the variation of RT position, we find that the intensity peaks of amplitude distributions and the central regions of phase distributions have strong connections with RT positions. Meanwhile, the boundaries of circular rings include structure information of the 3DFFDs. Hence, in the 3DFFD-based position inferring procedure, the suggested model should pay more attention to the aforestated parts. To this end, the gradient distributions of 3DFFDs are used to find these important regions and guide attention blocks to focus more on the features of intensity peaks, phase central regions, and boundaries, allowing for a more efficient and robust establishment of mapping relationships between the 3DFFDs and the incident angles.

TABLE I
CONFIGURATION OF THE PROPOSED GGAN

Attentional reconstruction network	Attentional localization network
Input: $K \times L \times W \times 2$	Input: $K \times L \times W \times 1$
Conv2D_1, $n = 128$, $ks = 7$	Conv2D_5, $n = 64$, $ks = 7$
Conv2D_2, $n = 128$, $ks = 5$	Conv2D_6, $n = 64$, $ks = 5$
Conv2D_3, $n = 256$, $ks = 5$	Conv2D_7, $n = 128$, $ks = 5$
Conv2D_4, $n = 512$, $ks = 7$	Conv2D_8, $n = 128$, $ks = 5$
TransConv2D_1, $n = 512$, $ks = 7$	LSTM, $n = 128$, $ks = 3$
TransConv2D_2, $n = 256$, $ks = 5$	Dense_1, $n^\star = 64$
TransConv2D_3, $n = 128$, $ks = 5$	Dense_2, $n^\star = 16$
TransConv2D_4, $n = 2$, $ks = 7$	Dense_3, $n^\star = 2$
Output: $K \times L \times W \times 2$	Output: 2×1

III. PROPOSED MODEL

The schematic of the proposed model is illustrated in Fig. 2. It consists of two subnetworks: 1) an attentional reconstruction network that consists of an encoder and a decoder that aim for reconstructing 3DFFDs from received signals and 2) an attentional localization network that is comprised of two encoders with the same structure and an MLP that are utilized to estimate the RT positions according to reconstructed results. The detailed configuration of GGAN is tabulated in Table I. As the structures of two encoders in the attentional localization network are consistent, the configuration of only one encoder is listed in Table I. K, L, and W refer to the slice number, length, and width of the reconstructed 3DFFD, respectively. n and ks denote the kernel number and kernel size of the convolutional layers, and $n^\star$ represents the number of densely connected neurons. All convolution strides in the proposed model are set to 1. The 3-D amplitude and phase distributions sparsely sampled by the cluster feed are considered the input

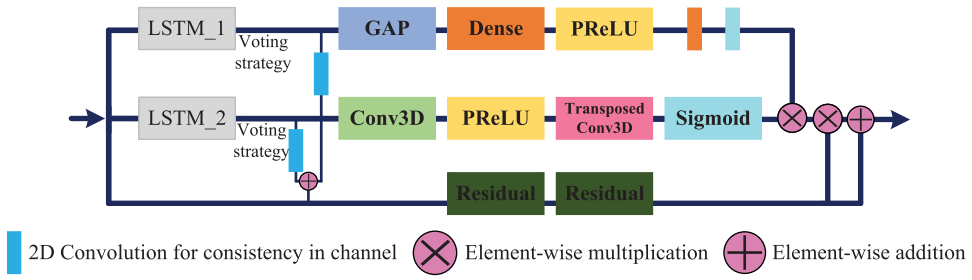

Fig. 3. Graphical illustration of the proposed LSTM-based attention block, which realizes spatial attention and channel attention simultaneously according to spatial features and dependencies of the 3DFFD.

of the attentional reconstruction network. In order to meet the requirement of the input dimension, the undersampled amplitude and phase distributions were padded with zeros and concatenated to create a two-channel input. The attentional reconstruction network simultaneously recovers features for both distributions, resulting in a two-channel output. Due to the high correlations between amplitude and phase distributions of the 3DFFDs, this proposed technique allows them to contribute supplemental information to each other in the distribution reconstruction stage. To explore a larger hypothetical space, we adopt two encoders in the localization network, which takes reconstructed amplitude and phase distribution as input, respectively. The 2-D convolutional and transposed convolutional layers are used to extract and recover features of 3DFFDs, respectively. Note that K convolutional or transposed convolutional layers are applied to K slices separately, which helps reserve the latent intraslices' relationships within 3DFFD. Each set of convolutional layers is followed by an LSTM-based attention block to establish the intraslices' dependences within 3DFFD and accomplish feature improvements. Gradient guidances are applied in the localization stage.

A. Attention Mechanism for 3DFFD Features

As analyzed in Section II, 3DFFDs under various RT position conditions are made up of salient regions and trivial parts. Furthermore, the thermal sky noise background, ground spillover noise, and receiver electronics noise all degrade 3DFFDs, potentially lowering the efficiency and performance of the localization model. Hence, the attention mechanism is employed in the proposed model to automatically multiply each 3DFFD feature in the hidden layers by a weight. Specifically, LSTM-based attention blocks and gradient guidances enable encoders to selectively emphasize representative features while weakening trivial ones. Apart from that, the attention blocks with feature recalibration capability achieve denoising and assist the decoder in synthesizing features.

1) LSTM-Based Attention Block: The LSTM module with gate layers is capable of selectively integrating features of FFD slices (the two-channel combination of amplitude and phase slices or either of them) depending on the importance of each slice's features [22], resulting in meaningful features and spatial relationships of FFD slices for each incident angle. Given the intermediate features of FFD slices, $\boldsymbol{a} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_K] \in \mathbb{R}^{K\times l\times w\times n}$, where K is the length of slice sequence. K, l, and w bound the spatial size. n represents the channel dimension. The spatial features on each channel are one type of representation of FFD slices. The LSTM module accepts $\boldsymbol{a}$ as an input and generates the hidden states $\boldsymbol{h} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K]$, where $\boldsymbol{h} \in \mathbb{R}^{K\times l\times w\times n'}$ and n' represents the channel dimension of the output after the LSTM module. The hidden state $\boldsymbol{h}_k$ refers to the representations and dependencies of the first k FFD slices. Instead of conventionally exploring the last hidden state $\boldsymbol{h}_K$, we introduced the voting strategy to incorporate all hidden states, giving our model richer dynamics. The voting strategy is given as follows:

$$\boldsymbol{v} = \text{softmax}(\boldsymbol{W}_v(E(\boldsymbol{h})) + \boldsymbol{b}_v). \tag{1}$$

E denotes the expectation for $\boldsymbol{h}$ across the last three axes, which results in a K dimensional vector. $\boldsymbol{W}_v \in \mathbb{R}^{K\times K}$ is the expectation-votes' parameter matrix, and $\boldsymbol{b}_v$ is the bias vector. The softmax function was utilized to normalize these votes to the $(0, 1)$ range. The integrated hidden state $\boldsymbol{h}_v$ is the weighted sum of $\boldsymbol{h}$ according to votes $\boldsymbol{v}$. By exploring $\boldsymbol{h}_v$, the channel attention and spatial attention were applied to adaptively search for the important channels in the last dimensionality and adaptively enhance spatial features in the remaining dimensionality.

Fig. 3 depicts the overview of our proposed LSTM-based attention block, which is comprised of three subbranches. In the top branch, the global average pooling was applied to the integrated hidden state $\boldsymbol{h}_{v1} \in \mathbb{R}^{l\times w\times n}$ with the same channel dimension as $\boldsymbol{a}$, yielding the global channel descriptor of $\boldsymbol{h}_{v1}$. Afterward, dense layers with parametric rectify linear unit (PReLU) [23] and sigmoid function were utilized to compress the channel descriptor to a quarter of its original length and then stretch it back, which allows mapping the channel descriptor to channel attention mask $\boldsymbol{M}_{ch}$.

The middle branch achieves spatial attention. We used the LSTM module with the voting strategy to produce $\boldsymbol{h}_{v2} \in \mathbb{R}^{l\times w\times K}$ with a size of $l \times w$ in spatial dimensions and K in the channel dimension. We reshaped $\boldsymbol{h}_{v2}$ to $\boldsymbol{h}'_{v2} \in \mathbb{R}^{K\times l\times w}$ with a size of $K \times l \times w$ in spatial dimensions, which is considered the raw attention mask. The 3-D convolutional and transposed convolutional layers with a kernel size of 3 and a PReLU and sigmoid function were exploited to excite the raw mask, resulting in the spatial attention mask $\boldsymbol{M}_{sp}$.

The remaining branch takes responsibility for FFD feature refining with the residual blocks [24] as the backbone. The 2-D convolutional layers with a kernel size of 1 were used to

make the channel dimensions of $\boldsymbol{h}_{v1}$, $\boldsymbol{h}_{v2}$, and $\boldsymbol{a}$ consistent. Thus, we could transfer the supplementary information of dimension-adjusted hidden states to the features of each slice of $\boldsymbol{a}$. Define the final output of the bottom branch as $\boldsymbol{a}_r$, and the improved features of FFD slices $\boldsymbol{a}_l$ can be calculated as follows:

$$\boldsymbol{a}_l = \boldsymbol{a}_r \otimes \left(\boldsymbol{M}_{ch} \otimes \boldsymbol{M}_{sp} + 1\right). \tag{2}$$

2) Gradient Guidance: As discussed in Section II, some parts of 3DFFDs provide a significant contribution to the localization task. The intensity peaks of amplitude distributions and the central regions of phase distributions of 3DFFDs are extremely sensitive to RT positions. The boundaries of concentric rings reflect essential structure information of 3DFFDs. Since the large gradient regions of 3DFFDs correspond to the aforementioned salient parts, we directly strengthened the intermediate features derived from 3DFFDs' large gradient regions without the need for data-driven training to make these representative parts have a great impact on final localization.

It is worth mentioning that, since the received signals are severely undersampled noisy data of 3DFFDs, gradient distributions reflect limited information about critical areas of the undersampled 3DFFDs. The reconstructed results, on the other hand, are of high quality, allowing for meaningful gradient distributions. Therefore, the gradient guidances are only applied in the attentional localization network.

For each RT's position, the gradient distribution of the 3DFFD is calculated by the Sobel operator as

$$\boldsymbol{G}_l = \hat{\boldsymbol{A}} * \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \ \boldsymbol{G}_w = \hat{\boldsymbol{A}} * \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}. \tag{3}$$

$\hat{\boldsymbol{A}} \in \mathbb{R}^{K \times L \times W}$ refers to the original 3DFFD (amplitude or phase distribution in this case) before fed into localization network. $*$ here denotes the convolution operation. $\boldsymbol{G}_l$ and $\boldsymbol{G}_w$ represent derivative approximations of $\hat{\boldsymbol{A}}$ in the length and width direction, respectively. In order to reduce the computational load, we adopted the sum of absolute values to approximate the gradient magnitude as $\boldsymbol{G} = |\boldsymbol{G}_l| + |\boldsymbol{G}_w|$.

Given the intermediate features of the 3DFFD, we can figure out the linkages of all features to different patches of the same size in the original 3DFFD based on the receptive field (RF) theory [25]. The RF size means that the feature is associated with an $RF \times RF$ region of 3DFFD. Since LSTM modules and residual blocks in attention blocks with the kernel size of 1 do not affect RF, RF of the feature in any hidden layers can be calculated by iterating the following equation using the initial value 1 for RF_0 and s_0:

$$RF_m = RF_{m-1} + (ks_m - 1) \prod_{j=0}^{m-1} s_j \tag{4}$$

where ks and s represent the kernel size and stride of the convolutional layer, respectively. The subscripts m and j refer to the mth and jth convolutional layers outside the attention blocks. The expectation of the gradients of the receptive $RF \times RF$ area is considered the essentiality score of the linked feature, which is then normalized by the highest score among all intermediate features. All normalized scores, which serve as the guidance mask $\boldsymbol{M}_g^a$ for the amplitude distribution, operate on $\boldsymbol{a}_l$ as follows:

$$\boldsymbol{a}_l^g = \boldsymbol{a}_l \otimes \left(\boldsymbol{M}_g^a + 1\right). \tag{5}$$

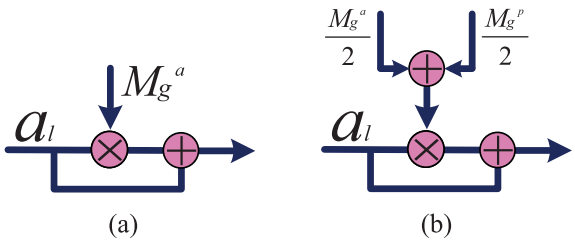

Fig. 4. Illustration of the gradient guidances. $\boldsymbol{M}_g^a$ and $\boldsymbol{M}_g^p$ denote the guidance masks derived from the gradient distribution of amplitude distribution and that of phase distribution, respectively. (a) Guidance for amplitude, (b) Guidance for phase.

The phase central regions of the 3DFFD have strong ties with the incident angle. Nevertheless, the corresponding intermediate features have low essentiality scores since the gradient distribution of the phase distribution could not identify the central regions. Consequently, based on the position consistency between phase central regions and amplitude peaks, we employed amplitude and phase guidance masks simultaneously for the intermediate features of the phase distribution. Fig. 4 shows the rule of gradient guidances.

B. Encoders for 3DFFDs

Encoders in reconstruction and localization networks aim at extracting informative features of 3DFFDs, which permits the reconstruction with authenticity and localization with high precision. Given the 3DFFD under a certain incident angle condition, the 2-D convolutional layers with learnable parameters activated by PReLUs were first implemented to search for latent features of the FFD slices. In consideration of the feature similarity between the FFD slices, the 2-D convolutional layer used by one FFD slice shares the parameters with other convolutional layers, which improves the computational efficiency. Moreover, the attention mechanism was applied to assign each intermediate feature a weight automatically to improve the robustness of the intermediate features of FFD slices. Meanwhile, the slices' spatial dependencies and distribution features generated by LSTM modules were incorporated into the intermediate features, which produces comprehensive and detailed representations of the 3DFFD. It is worth noting that, in the attentional localization network, we employed the LSTM with voting strategy as the final step of encoding to further extract the latent features and the dynamic intraslices' relationships.

C. Decoder and Position Estimator

The goal of the decoder is to recover 3DFFDs from the coded features. The decoder uses the 2-D transposed convolutional layers with shared and learnable parameters activated

by PReLUs to perform feature recovery of 3DFFDs in spatial dimensions. Meanwhile, skip connections were utilized between the encoder and the decoder to provide supplementary information for feature synthesizing. To achieve adaptive recalibration of the intermediate synthesized features, we also employed the LSTM-based attention block after the transposed convolutional layers throughout the reconstruction process.

The position estimation is accomplished using an MLP, which consists of three dense layers. Dense layers are activated by the PReLUs, except the last one that uses sigmoid transformation. The MLP takes the combination of features and dependencies of 3DFFDs extracted by two encoders as the input and regresses the coded information to RT positions.

D. Loss Functions

By minimizing loss functions, i.e., the distances between predicted results and the labels, we can optimize the learnable parameters in the proposed model. As a result, the proposed model provides a competitive performance on reconstruction and localization tasks.

The reconstruction network was trained with the mean square error (mse) loss ℓ_{mse}, which was exploited to constrain the average differences between the reconstructed results and true 3DFFDs in pixel lever. However, if solely focusing on pixel lever similarity, the irregular abrupt changes caused by the 2π phase jumps in phase distributions could make reconstruction problematic. Consequently, we additionally took advantage of the perceptual consistency loss ℓ_{P} [26] to boost the reconstruction performance. The total loss ℓ_{Rec} of the reconstructed result is calculated as

$$\ell_{\text{Rec}} = \ell_{\text{mse}} + \ell_{\text{P}} \tag{6}$$

where

$$\ell_{\text{mse}} = \mu_1 \frac{1}{\text{KLW}} \left\| \boldsymbol{A}^a - \hat{\boldsymbol{A}}^a \right\|_F^2 + \mu_2 \frac{1}{\text{KLW}} \left\| \boldsymbol{A}^p - \hat{\boldsymbol{A}}^p \right\|_F^2 \tag{7}$$

and

$$\ell_{\text{P}} = \frac{\mu_3}{K} \sum_{\upsilon=1}^{V} \frac{1}{l_\upsilon w_\upsilon n_\upsilon} \left\| g_\upsilon\left(\boldsymbol{A}^p\right) - g_\upsilon\left(\hat{\boldsymbol{A}}^p\right) \right\|_F^2 \tag{8}$$

where $\boldsymbol{A}$ and $\hat{\boldsymbol{A}} \in \mathbb{R}^{K\times L\times W}$ denote the true label and the reconstructed result, respectively, and superscripts a and p represent amplitude and phase distribution, respectively. $\|\cdot\|_F$ refers to the Frobenius norm. In (8), $g_\upsilon(\boldsymbol{A}^p)$ signifies abstract features of the phase distribution extracted by the υth hidden layer segment of the pretrained VGG19 [27], which are considered an overall perception of the phase distribution. Specifically, three segments of VGG19 are adopted, which are the first convolution block, the first two convolution blocks, and the first three convolution blocks of VGG19. To satisfy the input dimension requirement of VGG19, we repeated the FFD slices three times along the channel dimension and fed the FFD slices into VGG19 in sequence, resulting in abstract features $\in \mathbb{R}^{K\times l_\upsilon \times w_\upsilon \times n_\upsilon}$, where K refers to slice number, and l_υ, w_υ, and n_υ bound the size of output features of the υth segment. μ_1, μ_2, and μ_3 govern the effect of each part of the loss function on parameter optimizing.

In order to achieve competitive performance for the localization network, we aim at minimizing the integration of square errors between the predicted angles and the angle labels, which is calculated as follows:

$$\ell_{\text{Loc}} = \mu_4\left(\theta - \hat{\theta}\right)^2 + \mu_5\left(\phi - \hat{\phi}\right)^2 \tag{9}$$

where θ and ϕ denote the zenith and azimuth angles, respectively, and $\hat{\theta}$ and $\hat{\phi}$ represent corresponding predicted results. μ_4 and μ_5 control the effect of each angle loss on parameter optimizing.

IV. Experiments

A. Dataset and Evaluation Metrics

To facilitate the analysis, a scaled version of the FAST with a 30-m-diameter aperture and a focal length to diameter ratio of 0.467 using the 19-beam L-band receiver is selected for this study [28]. FEKO 7.0 was used to simulate the 3DFFDs near the reflector focus with various incident angles ranging from 0° to 4° in zenith with the interval of 0.1° and 0°–360° in azimuth with the interval of 5° at 1.25 GHz. The received signals, i.e., undersampled 3DFFDs, were obtained by the 19-element cluster feed sampling field distributions in the focal plane and vicinity planes shifted up and down by λ. To evaluate the robustness of the proposed model against noisy data, complex Gaussian noise was added to undersampled 3DFFDs with the signal-to-noise ratios (SNRs) ranging from -6 to 15 dB with a 3-dB interval. The corresponding reconstruction labels are $5\times 21\times 21$ sampled pixels of field distributions with $\lambda/2$ interval between the adjacent slices and 0.054-m sampling interval along other axes. To facilitate the model's understanding of the data, we normalized the amplitude distributions by the maximum values and phase distributions by 2π. The randomly shuffled dataset was split: 85% for training and 15% for testing.

To evaluate the authenticity of the reconstructed 3DFFDs, we adopt two of the most commonly used quantitative metrics in the image reconstruction field: the peak SNR (PSNR) and structural similarity (SSIM).

The mean angle (MA) between the true and predicted direction vectors is used to evaluate the position estimation performance, which is calculated as

$$\text{MA} = \frac{1}{N}\sum_{i=1}^{N} \angle\left(\boldsymbol{\zeta}_i, \hat{\boldsymbol{\zeta}}_i\right) \tag{10}$$

where

$$\begin{aligned}\angle\left(\boldsymbol{\zeta}_i, \hat{\boldsymbol{\zeta}}_i\right) &= \arccos\left(\frac{\boldsymbol{\zeta}_i \hat{\boldsymbol{\zeta}}_i}{|\boldsymbol{\zeta}_i||\hat{\boldsymbol{\zeta}}_i|}\right) \\ &= \arccos\Big(\sin\theta_i\cos\phi_i\sin\hat{\theta}_i\cos\hat{\phi}_i + \sin\theta_i\sin\phi_i \\ &\quad \times \sin\hat{\theta}_i\sin\hat{\phi}_i + \cos\theta_i\cos\hat{\theta}_{\text{i}}\Big)\end{aligned} \tag{11}$$

where $\boldsymbol{\zeta}_i$ refers to the unit vector in the direction of true RT position and $\hat{\boldsymbol{\zeta}}_i$ is the corresponding prediction with our model. N denotes the count of test samples. For each nonoverlapping beam of the scaled FAST model, the positions of detected RTs can be approximately limited to the angular region bounded

TABLE II
RECONSTRUCTION PERFORMANCE OF THE PROPOSED METHOD AND COUNTERPARTS

Reconstruct separately		Reconstruct with MSE loss		Proposed method	
PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
31.91	0.89	32.40	0.91	33.14	0.92
19.46	0.61	23.20	0.79	25.20	0.82

by the half-power beam solid angle [29], [30]. We neglect the slight asymmetry of beams and differences between the half-power beam widths (HPBWs) of beams caused by the Coma effect and consider the HPBW of the central beam as the angular resolution of the telescope system. In this study, the HPBW of the central beam is 0.554°, which was obtained by GRASP 10.3.0 under the condition of full Gaussian illumination of the aperture. Hence, when choosing a position in the angular region as the localization result, the lower bound of the localization error is 0°, and the upper bound is 0.554°. To facilitate the comparison with the proposed method, we select 0.277° as the reference value of the localization error of the scaled FAST model.

FOV refers to the instantaneously observable region of the radio telescope. Since the HPBWs of adjacent beams of the scaled FAST model do not overlap, the FOV could be calculated as FOV$= N_f\Omega$ with ignoring small variations within beams. N_f denotes the number of feed elements. Ω is the half-power beam solid angle of the central beam, which is equal to $\int_0^{360\pi/180}\int_0^{0.277\pi/180}\sin\alpha\,\mathrm{d}\alpha\,\mathrm{d}\beta$ steradians. The FOV of the scaled FAST model is about 1.40×10^{-3} steradians and 4.58 in square degrees.

B. Training Procedure

First, the attentional reconstruction network was optimized by minimizing ℓ_{Rec} based on the minibatch stochastic gradient descent, which was performed by the Adam optimizer with the default setting [31]. We trained the model with a minibatch size of 32, and an initial learning rate of 1×10^{-4} decreased by a factor of 10 every 40 epochs until model convergence was reached.

In the next stage, the localization network was attached to the reconstruction network. The integrated network was also trained using Adam optimizer by minimizing ℓ_{Loc} with a minibatch size of 32 and an initial learning rate of 1×10^{-5} that declined tenfold every 40 epochs. Instead of fixing the learned parameters in the first stage, we fine-tuned these parameters at a relatively lower learning rate, which makes our model more flexible. In both stages, we dynamically regulated μ in (7)–(9) to ensure that each component of the loss functions contributes equally to the optimization process.

C. Reconstruction Performance

Table II compares the average reconstruction performance of the proposed reconstruction method and the counterparts (i.e., reconstructing amplitude and phase separately, and reconstructing only with mse loss) on the test dataset. Rows 3 and 4 correspond to amplitude and phase distributions, respectively. The reconstruction results of amplitude distributions outperform those of the phase distributions. This is due to that the phase distributions do not have the same intensity tapering property as amplitude distributions. Thus, drastically undersampled phase distributions contain relatively less pattern information. As can be seen from the first and third columns, the synchronous reconstruction method achieves better performance compared with the separate reconstruction method, especially for phase distributions that have a big jump in authenticity. It demonstrates that the synchronous reconstruction strategy can leverage the knowledge used for one distribution reconstruction to assist the other distribution reconstruction. Moreover, it can be seen from the last two columns that the supervision of perceptual consistency can improve not only the reconstruction performance of phase distributions but also that of amplitude distributions by a small margin.

Fig. 5 presents the examples of reconstruction under various RT position conditions at SNR $=$ 15 dB. By comparing the ground truths and reconstructed results of amplitude distributions, we can see that the distortions of the amplitude distributions mainly occur in the side lobes. Besides, Fig. 5(c) indicates that the sidelobe with extremely low intensity may not be reconstructed. For the reconstructed phase distributions, the distortions are manifested as the introduction of noise signals in relatively flat regions. We can see that, despite minor defects occurring in the reconstructed distributions, they do not give a significant impact on the reconstruction results of both the amplitude and phase distributions under various incident angle conditions. Thus, these amplitude and phase data can make the localization robust in the second stage. In addition, the results show strong correlations between the central regions of FFD slices and the RT positions. With increasing the zenith angle, the central regions of both amplitude and phase distributions deviate from the slices. When the azimuth angle changes, the amplitude and phase distributions are rotated accordingly.

D. Localization Performance

1) Effects of FFD Slice Numbers: Table III compares the average localization performance on the test dataset with different numbers of reconstructed FFD slices. "1 slice" represents the slice on the focal plane. "3 slices" refers to slices on the focal plane and λ offset planes (toward and away from the reflector). "5 slices" has one more slice on $\lambda/2$ offset plane (toward the reflector) than "4 slices". By analogy, $\lambda/4$ offset plane (away from the reflector) for "6 slices", and $\lambda/4$ offset plane (toward the reflector) for "7 slices." Results show that the localization using merely a focal plane slice has the worst performance. The localization performance could be considerably improved by utilizing more FFD slices, which contains more spatial features and dependencies, whereas the performance has substantially remained unchanged when using more than five FFD slices. This is attributed to the fact that the received signals carry a finite amount of information regarding 3DFFDs, which determines the upper bound of

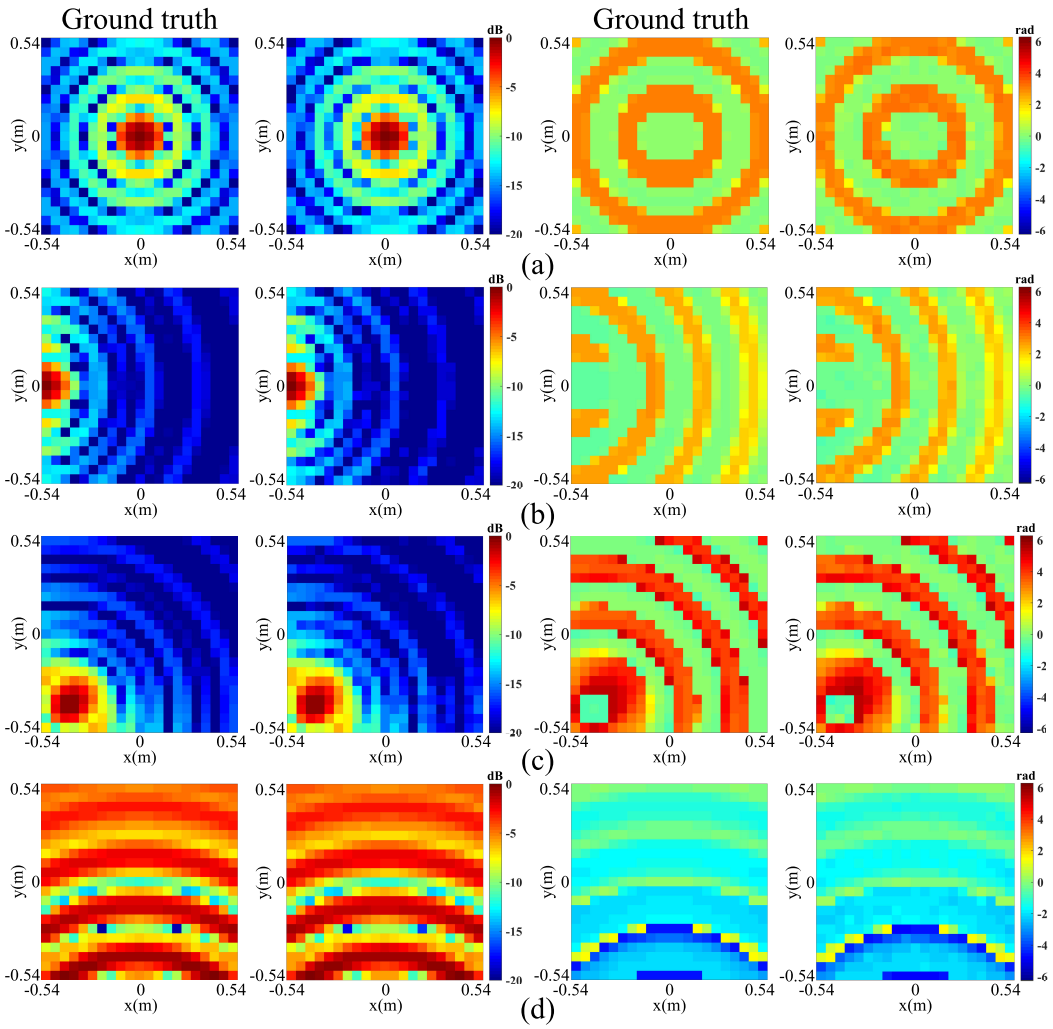

Fig. 5. Examples of reconstruction. The first two columns are amplitude slices, and the others are phase slices. (a) corresponds to the FFD slice on the focal plane under $\theta = 0°$, $\phi = 0°$ condition. (b) Focal plane, $\theta = 2°$, $\phi = 0°$. (c) $\lambda/2$ offset plane (away from the reflector), $\theta = 2°$, $\phi = 45°$. (d) λ offset plane (away from the reflector), $\theta = 4°$, $\phi = 90°$.

TABLE III

MEAN LOCALIZATION ERRORS WITH THE VARIATION OF SLICE NUMBERS

1 slice	3 slices	4 slices	5 slices	6 slices	7 slices
0.285°	0.254°	0.237°	0.223°	0.222°	0.225°

TABLE IV

MEAN LOCALIZATION ERRORS WITH ATTENTION BLOCK ABLATIONS IN DIFFERENT PARTS OF GGAN

Encoder[1]	Encoders[2]	Decoder	Coherent attention
0.239°	0.244°	0.241°	0.223°

localization accuracy. In order to release the computational burden, the number of slices was set to 5 in this research.

2) Ablation Experiments: To demonstrate the effectiveness of the proposed LSTM-based attention block, we investigated the effects of attention block ablations on the localization performance of the proposed GGAN. We retained the feature refining branch and only wiped off attention branches (i.e., $\boldsymbol{a}_l = \boldsymbol{a}_r$). The obtained results are tabulated in Table IV, where Encoder[1], Encoders[2], and Decoder represent the removal of all attention blocks in the encoder of the reconstruction network, that in encoders of localization network, and that in the decoder, respectively. As can be seen from Tabel IV, no matter where the attention blocks are removed, the localization performance suffers accordingly, indicating that the proposed attention block could assist in feature coding and synthesizing.

3) Comparisons With Attention-Based Methods: We compared the localization performance of the proposed GGAN and LSTM-based attentional network (LSTM-AN, i.e., GGAN without gradient guidances) with two attention-replaced networks. These two networks are the results of replacing all the attention blocks of LSTM-AN with two competing attention strategies: 1) the convolutional block attention module (CBAM) [16] that employs channel attention and spatial attention sequentially based on the intermediate features and 2) squeeze-and-excitation (SE) block [18] that aims at achieving attention in the channel domain by exploring the channel

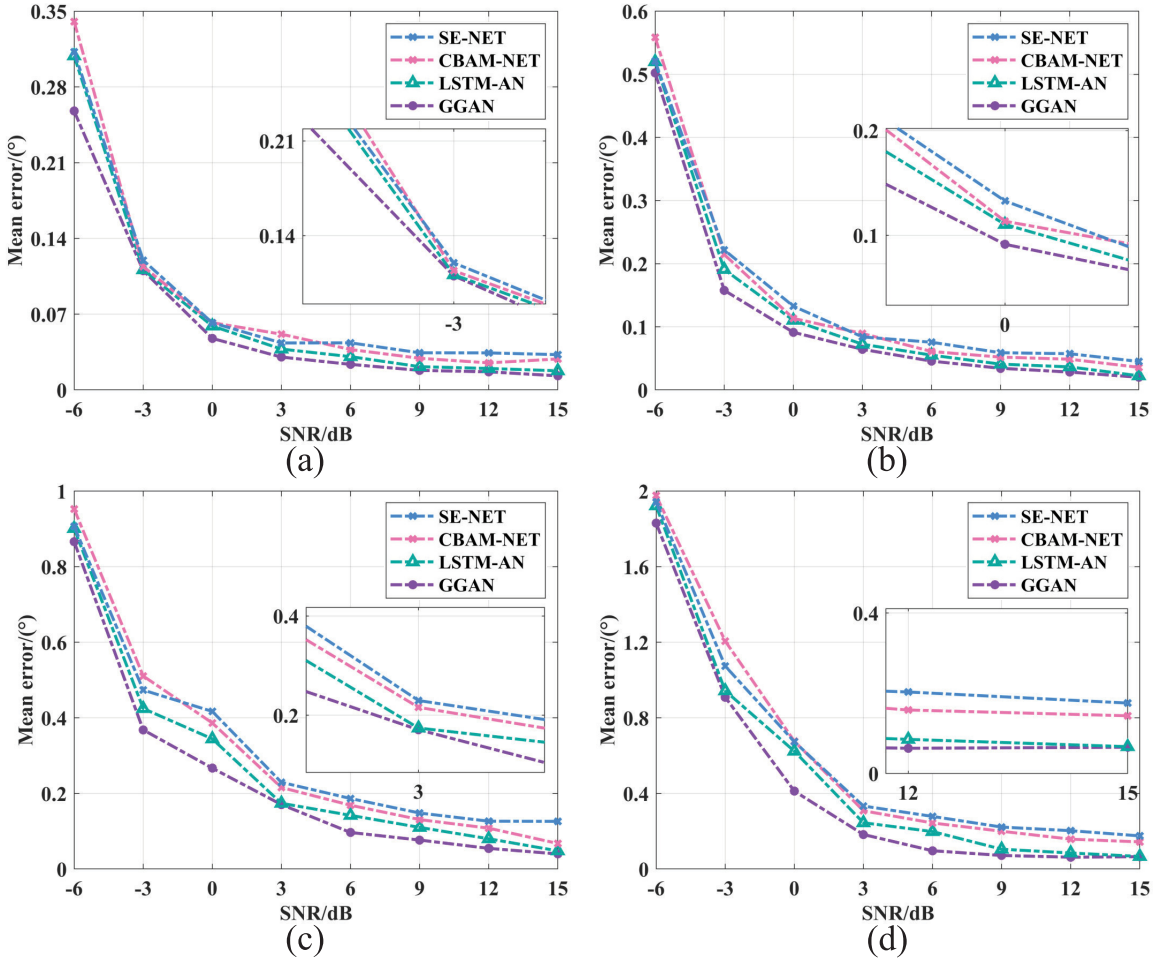

Fig. 6. Localization performance of attention-based methods with the variation of SNRs. (a) corresponds to region 1. (b) corresponds to region 2. (c) corresponds to region 3. (d) corresponds to region 4.

descriptors. To fairly compare those methods, CBAM and SE blocks were equipped with the same feature refining branch as our proposed attention block. In other words, $\boldsymbol{a}_r$ was elementwisely multiplied by the attention masks generated by the competing attention methods. To provide qualitative analysis, we divided RT positions into four regions according to the central regions of 3DFFDs: region 1 with zenith angles between 0° and 1° (central regions in the reconstructed slices), region 2 with zenith angles between 1° and 2° (slightly out of the slices), region 3 with zenith angles between 2° and 3° (out of the slices), and region 4 with zenith angles between 3° and 4° (farthest from the slices).

Fig. 6 shows the localization performance curves concerning SNRs. As we can see, the average localization errors of all methods decrease sharply at low SNRs and decrease relatively slowly when the SNRs are above 6 dB. Meanwhile, the average localization errors increase significantly as the zenith angle increases, which demonstrates that the central regions of 3DFFDs play an important role in the localization task. Instead of only exploring spatial features of 3DFFDs to achieve attention inference, our methods also consider spatial relationships within 3DFFDs. Hence, LSTM-AN and GGAN achieve more robust performance against noisy data compared with the CBAM network (CBAM-NET) and the SE network (SE-NET). Moreover, since the enhancement of the salient central regions and boundaries of 3DFFDs, the gradient guidances further lower localization errors. Compared with the radio telescope's localization error (i.e., 0.277° in this experiment), the proposed GGAN could achieve lower localization error in regions 1 and 2 at SNRs greater than -3 dB and in regions 3 and 4 at SNRs greater than 3 dB, which indicates excellent localization performance of the proposed GGAN method when the zenith angle ranges between 0° and 2° and satisfactory localization performance under relatively high SNR condition when the zenith angle ranges between 2° and 4°. Since the GGAN method achieves better localization accuracy compared with the inherent localization accuracy of the scaled FAST model when the zenith angle ranges between 0° and 4° and the azimuth angle ranges between 0° and 360°, the FOV of the GGAN method is $\int_0^{360\pi/180}\int_0^{4\pi/180}\sin\alpha\,\mathrm{d}\alpha\,\mathrm{d}\beta$ steradians, which approximately is 50.25 in square degrees, a 10.97 times improvement than the radio telescope's FOV.

V. Conclusion

This article presents a GGAN to improve localization accuracy for RTs by exploring 3DFFDs. The first component of the proposed GGAN is the attentional reconstruction network, which can be used to construct higher-resolution 3DFFDs from undersampled 3DFFDs. The attentional localization network, as the second part, aims at estimating RTs' positions according to the reconstructed results. The attention block in GGAN enables attention inference based on spatial features and intraslices' relationships of 3DFFDs generated by the LSTM with the voting strategy module. The gradient distributions of reconstructed 3DFFDs provide a guide for the attention block in the localization network without the need for data-driven training. This allows the salient elements of 3DFFDs to significantly influence the prediction of the RT's position. Experiments have shown that the proposed model not only outperforms the radio telescope's angular resolution but also broadens the FOV of the radio telescope.

References

[1] Z. Zhang, K. Yan, C. Li, G. Zhang, and F. Wang, "Intergalactic medium dispersion measures of fast radio bursts estimated from IllustrisTNG simulation and their cosmological applications," *Astrophys. J.*, vol. 906, no. 1, pp. 49–58, Jan. 2021.

[2] A. Walters, A. Weltman, B. Gaensler, Y. Ma, and A. Witzemann, "Future cosmological constraints from fast radio bursts," *Astrophys. J.*, vol. 856, no. 1, pp. 65–72, Mar. 2018.

[3] C. Law *et al.*, "*realfast*: Real-time, commensal fast transient surveys with the very large array," *Astrophys. J. Suppl. Ser.*, vol. 236, no. 1, pp. 8–15, May 2018.

[4] E. Fomalont and M. Wright, "Interferometry, aperture synthesis, and VLBI," in *The WSPC Handbook of Astronomical Instrumentation: Radio Astronomical Instrumentation*, vol. 1, A. Wolszczan, Ed. Singapore: World Sci., 2021, pp. 25–47.

[5] P. Thavasimani and A. Scaife, "Square kilometre array : Processing voluminous MeerKAT data on IRIS," 2021, *arXiv:2105.14613*.

[6] A. R. Thompson, J. M. Moran, and G. W. Swenson, "Response of the receiving system," in *Interferometry and Synthesis in Radio Astronomy*, 3rd ed. Cham, Switzerland: Springer, 2017, pp. 207–254.

[7] C. J. Salter, "Single-dish radio telescopes," in *The WSPC Handbook Astronomical Instrumentation: Radio Astronomical Instrumentation*, vol. 1, A. Wolszczan, Ed. Singapore: World Sci., 2021, pp. 1–24.

[8] K. F. Warnick, R. Maaskant, M. V. Ivashina, D. B. Davidson, and B. D. Jeffs, *Phased Arrays for Radio Astronomy, Remote Sensing, and Satellite Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2018.

[9] G. H. Hilmarsson, "Targeted fast radio burst searches with the Efflesberg 100-m radio telescope," Ph.D. dissertation, Univ. Bonn, Bonn, Germany, 2021. [Online]. Available: https://hdl.handle.net/20.500.11811/8999

[10] K. F. Warnick, R. Maaskant, M. V. Ivashina, D. B. Davidson, and B. D. Jeffs, "High-sensitivity phased array receivers for radio astronomy," *Proc. IEEE*, vol. 104, no. 3, pp. 607–622, Mar. 2016.

[11] S. Osłowski *et al.*, "Commensal discovery of four fast radio bursts during Parkes pulsar timing array observations," *Monthly Notices Roy. Astronomical Soc.*, vol. 488, no. 1, pp. 868–875, Jun. 2019.

[12] P. Jiang *et al.*, "Commissioning progress of the FAST," *Sci. China Phys., Mech. Astron.*, vol. 62, no. 5, pp. 959502–959523, Mar. 2019.

[13] D. Wu *et al.*, "Fast searches for astronomical transients based on focal-field feature matching," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 3, pp. 482–486, Mar. 2019.

[14] D. Wu, H. Cao, N. Lv, J. Fan, X. Tan, and S. Yang, "Feature matching conditional GAN for fast radio burst localization with cluster-fed telescope," *Astrophys. J.*, vol. 887, no. 1, pp. L10–L16, Dec. 2019.

[15] F. Wang *et al.*, "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3156–3164.

[16] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 3–19.

[17] J. Fu *et al.*, "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3146–3154.

[18] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.

[19] X. Xu *et al.*, "Asymmetric multi-task attention network for prostate bed segmentation in computed tomography images," *Med. Image Anal.*, vol. 72, pp. 102116–102126, Aug. 2021.

[20] X. Ouyang *et al.*, "Dual-sampling attention network for diagnosis of COVID-19 from community acquired pneumonia," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2595–2605, Aug. 2020.

[21] M. Ivashina and C. V. Klooster, "Focal fields in reflector antennas and associated array feed synthesis for high efficiency multi-beam performances," *TIJDSCHRIFT-NERG*, vol. 68, no. 1, pp. 11–19, Jan. 2003.

[22] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space Odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2016.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jul. 2015, pp. 1026–1034.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Feb. 2016, pp. 770–778.

[25] H. Le and A. Borji, "What are the receptive, effective receptive, and projective fields of neurons in convolutional neural networks?" 2017, *arXiv:1705.07049*.

[26] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 694–711.

[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, May 2015, pp. 1–14.

[28] S. L. Smith *et al.*, "Performance validation of the 19-element multibeam feed for the five-hundred-metre aperture spherical radio telescope," in *Proc. IEEE Int. Symp. Antennas Propag. USNC/URSI Nat. Radio Sci. Meeting*, Jul. 2017, pp. 2137–2138.

[29] J. K. Swiggum and P. A. Gentile, "On-the-fly mapping of new pulsars," *Astronom. J.*, vol. 156, no. 5, pp. 1–6, 2018.

[30] E. Petroff, J. W. T. Hessels, and D. R. Lorimer, "Fast radio bursts," *Astronom. Astrophys. Rev.*, vol. 27, no. 1, pp. 1–75, 2019.

[31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, May 2015, pp. 1–15.

Hailin Cao received the B.E. degree in electronic engineering and the M.S. and Ph.D. degrees in circuits and systems from Chongqing University, Chongqing, China, in 2002, 2006, and 2010, respectively.

He is currently a Professor with the Chongqing Key Laboratory of Space Information Network and Intelligent Information Fusion and the State Key Laboratory of Power Transmission Equipment and System Security and New Technology, Chongqing University. His research interests include sensor networks, large structural health monitoring technology, and antenna and array signal processing.

Junhui Peng received the B.E. degree in communication engineering from Nanchang University, Nanchang, Jiangxi, China, in 2019. He is currently pursuing the Ph.D. degree with the Chongqing Key Laboratory of Space Information Network and Intelligent Information Fusion, Chongqing University, Chongqing, China.

His research is focused on deep learning and array signal processing.

Jin Fan received the B.Sc. degree in electrical and information engineering from the University of Science and Technology of China, Hefei, China, in 2007, the M.Sc. degree in communication and information engineering from the Graduate University of Chinese Academy of Sciences, Beijing, China, in 2010, and the Ph.D. degree in astronomical techniques and methodology from the University of Chinese Academy of Sciences, Beijing, in 2018.

From 2010 to 2019, she was with the Department of FAST, Chinese Academy of Sciences, National Astronomical Observatories of China, Beijing. Since April 2020, she has been a Senior Engineer with the Ming'antu Observing Station, Chinese Academy of Sciences, National Astronomical Observatories of China. She is also a Visiting Researcher with Brigham Young University, Provo, UT, USA, and the Chalmers University of Technology, Gothenburg, Sweden. Her research interests include ultrawideband (UWB) antennas and UWB feeds for reflector antennas, phased arrays, and phased array feed (PAF) technology, as well as high-sensitivity receiver systems for radio astronomical applications.

Wai Yan (Warren) Yong (Graduate Student Member, IEEE) received the bachelor's degree (Hons.) in electronic engineering from the Universiti Tun Hussein Onn Malaysia (UTHM), Parit Raja, Malaysia, in 2016, and the M.Phil. degree in electrical engineering from Universiti Teknologi Malaysia (UTM), Johor Bahru, Malaysia, in 2018. He is currently pursuing the D.Phil. degree with the Department of Radio System Engineering, University of Twente, Enschede, The Netherlands.

He is currently the Marie Curie Researcher funded by the European Union through the Marie Currie Horizon 2020 Project. He is also a Visiting Researcher with Gapwaves AB, Gothenburg, Sweden. His research interests include, but not limited to, millimeter-wave and THz antenna array designs, frequency-selective surfaces, phased array transceivers, and new materials for antenna and microwave devices.

Mr. Yong was a recipient of the UTHM Chancellor's Award during the 16th UTHM Convocation and the Best Postgraduate Student Award for his master's degree during the 60th UTM Convocation.

Decheng Wu received the B.E. degree in automation engineering from Southwest University, Chongqing, China, in 2012, and the Ph.D. degree in circuits and systems from Chongqing University, Chongqing, in 2020.

He is currently a Lecturer with the School of Automation, Chongqing University of Posts and Telecommunications, Chongqing. His research interests include deep learning, signal processing, and remote sensing.